DOCUMENT RESUME

ED 160 664 95 TM 007 984

AUTHOR Smith, Nick L.
TITLE The Limitations of ERIC in Reviewing Evaluation Literature. Paper and Report Series, No. 2. Research and Evaluation Program.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jun 78
NOTE 24p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Bibliographies; *Data Bases; *Evaluation; Evaluation Methods; *Indexing; *Information Retrieval; *On Line Systems; Program Evaluation; *Relevance (Information Retrieval)
IDENTIFIERS *ERIC

ABSTRACT

The Educational Resources Information Center (ERIC) system was used to identify 1,734 documents on program evaluation methodology. Attempts to further characterize these documents by subject were only partially successful because of the lack of sufficient depth of indexing and design constraints of the computer system. Attempts to characterize these documents by type of publication were also unsuccessful because so many of the documents had not been assigned publication type codes. When compared with a 1976 evaluation bibliography prepared by the Evaluation Center at Western Michigan University and edited by M.A. Bunda, only 27 percent of the bibliography's references were also listed in ERIC. Further investigation revealed that 28 percent of the bibliography was published or written before ERIC was incorporating such documents into its system, 22 percent were books or chapters of books (not usually made available through ERIC), nine percent were journal articles, and 14 percent were fugitive documents, which should generally be included in ERIC. Additional problems associated with non-hierarchical searching, system constraints on assignments of descriptors, and indexer limitations resulted in very little overlap between the ERIC search results and the Western Michigan bibliography. Researchers should be aware of the limitations of a search of the ERIC system. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160664

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

paper and report series

Research on Evaluation Program

Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204
Telephone: (503) 248-6800

IM007 984

No. 2 THE LIMITATIONS OF ERIC IN REVIEWING EVALUATION LITERATURE

NICK L. SMITH

June 1978

Nick L. Smith, Director
Research on Evaluation Program
Northwest Regional Educational Laboratory
710 S.W. Second Avenue, Portland, Oregon 97204

Published by the Northwest Regional Educational Laboratory, a private nonprofit corporation. The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

PREFACE

The Research on Evaluation Program is a Northwest Regional Educational Laboratory project of research, development, testing, and training designed to create new evaluation methodologies for use in education. This document is one of a series of papers and reports produced by program staff, visiting scholars, adjunct scholars, and project collaborators--all members of a cooperative network of colleagues working on the development of new methodologies.

There are 21,675 documents in ERIC descriptored "Evaluation". Can ERIC system features be used to pre-sort and categorize this literature by computer? How well does ERIC cover the evaluation literature? This report describes ERIC limitations which prevent computerized pre-screening of literature, and recounts the failure of ERIC to reconstruct an existing evaluation bibliography (only 1.2 percent of the items were identified by ERIC), casting serious doubt on the utility of ERIC searches of the evaluation literature.

Nick L. Smith, Editor
Paper and Report Series

THE LIMITATIONS OF ERIC IN REVIEWING EVALUATION LITERATURE[1]

It has become a common observation that educational researchers now produce studies and reports at a much faster rate than can be read and understood, even by colleagues. The engulfing literature long ago surpassed the individual researcher's capacity to absorb nonintegrated findings and as a profession "...we find ourselves in the mildly embarrassing position of knowing less than we have proven." (Glass, 1976, p. 8)

We recognize with Glass that "...the armchair literature review in which one cites a couple dozen studies from the obvious journals can't do justice to the voluminous literature of educational research that we now confront." (1976, p. 4) He and others (e.g., Light and Smith, 1971) have therefore stressed the need for integrative, scholarly reviews and have suggested procedures for combining and interpreting the findings from multiple studies. While these authors have dealt with the complex problems of integrating convergent data, I am concerned here with the more mundane problem of how we initially identify which studies to review. The old method of screening several hundred studies by hand is no longer feasible when the potentially relevant literature contains thousands, not hundreds, of citations. For example, we recently began a review of the literature on the methodology of program evaluation and found that in the ERIC (Educational Resources Information Center) system:

21,675 documents were descriptored "Evaluation,"

13,433 documents were descriptored "Program Evaluation,"

and 6,689 documents were descriptored "Evaluation Methods."

Since there were obviously too many documents to review by hand, we tried to use certain features of the ERIC system to reduce the collection.

Using the ERIC interactive computer search capability, we attempted to identify and characterize the literature of interest in our search.[2] Our purpose was to isolate the "core" literature and to reduce the number of citations to a manageable size of about 2,000. At that point we could print and screen 2,000 abstracts by hand, selecting a lesser number of documents for actual study. As a result of our efforts, we were able to make some general assessments of the limitations of the ERIC system in conducting such searches.

Characterizing Literature Using ERIC

One of the biggest problems in the early stages of an integrative literature review is to characterize the general nature of the relevant literature. This is a crucial step, since subsequent screening and review decisions are based, in part, on assumptions about the nature of the documents one is dealing with.

After isolating what appeared to be the relevant documents, we attempted to use ERIC to characterize their nature. Each document in ERIC is labeled by multiple descriptor words which can be used to build document collections. Each document is given a maximum of five major descriptors and as many minor descriptors as necessary (up through 1972, documents were given an average of 11 descriptors each (ERIC, 1974, p. 187)).

To build an initial document collection we first combined all those documents related to program evaluation, broadly construed, by using the following descriptors:

Program Evaluation	(13,433 documents)
Curriculum Evaluation	(2,139)
Formative Evaluation	(680)
Summative Evaluation	(595)
Medical Evaluation	(398)
Medical Care Evaluation	(34)
Dental Evaluation	(27)

This resulted in a total of 16,206 separate documents. In order to identify only the "core" or most significant program evaluation documents, this collection was restricted to only those using the above terms as major descriptors, reducing the collection to 9,179 documents.

Next we identified all documents related to methodology using the following descriptors:

Evaluation Methods	(6,689)
Research Methodology	(6,468)
Methods	(1,474)
Scientific Methodology	(492)
Inductive Methods	(478)
Methods Research	(344)
Deductive Methods	(272)

This collection totaled 15,506 documents (including both major and minor descriptors).

In order to identify those documents related to the methodology of program evaluation, we then crossed the 9,179 program evaluation documents with the 15,506 methodology documents and thus constructed a core collection of 1,734 documents with major descriptors of program evaluation and major or minor descriptors of methodology. We were successful, therefore, in isolating a core collection of documents that numbered less than 2,000.

1,734 abstracts are still a lot to review so we tried to use the ERIC system to characterize the nature of this core collection, It would have been most useful if the ERIC system could have simply printed a frequency list of the descriptors used in this core collection, but the system has no such feature. Instead, it was necessary to choose likely terms and inquire how many of the core collection used such terms as descriptors. Through such a strategy the following Table 1 was constructed.

Insert Table 1 about here

Table 1 entries total to 1,948 and not 1,734, since the subcategories are not mutually exclusive. Table 2 indicates the overlap between these subcategories.

Insert Table 2 about here

Notice that the two largest subcategories, "Data" and "Measurement" overlap considerably, (208 (23.4%) of the total 887 documents are in both categories) while "Design" and "Policy", among the smallest subcategories, hardly overlap at all (2 (1.6%) of the total 124 documents are shared).

Table 1
Subcategorization of Program Evaluation Methodology
Core Collection

Subcategory Descriptors	Number of Documents
1. Policy Legislation	55
2. Theories Models	241
3. Design Research Design Design Needs Design Preferences	71
4. Data Data Collection Data Processing Electronic Data Processing Statistical Data Analysis	407
5. Measurement Educational Assessment Needs Assessment Test	688
6. Skills Ethics Standards Qualifications Credentials Job Skills Research Skills	43
7. Training	112
8. Management Program Administration Administration Administrative Principles Administrator Characteristics Administrator Guides Administrative Responsibility Administrator's Role	128
9. Problems Needs	203

Table 2
Number of Documents Sharing Two Subcategory Descriptors

Subcategory	1. "Policy" (55)	2. "Theories" (241)	3. "Design" (71)	4. "Data" (407)	5. "Measurement" (688)	6. "Skills" (43)	7. "Training" (112)	8. "Management" (128)	9. "Problems" (203)
1. "Policy" (55)	--	6	2	12	15	2	3	6	9
2. "Theories" (241)		--	9	56	80	6	7	12	36
3. "Design" (71)			--	25	38	3	4	3	18
4. "Data" (407)				--	208	8	26	28	45
5. "Measurement" (688)					--	15	43	46	80
6. "Skills" (43)						--	6	3	7
7. "Training" (112)							--	8	11
8. "Management" (128)								--	11
9. "Problems" (203)									--

Only 1,245 documents of the total 1,734 in the core collection of program evaluation methodology can be placed in one or more of the subcategories listed in Table 1, leaving 489 documents, or 28 percent of the collection, uncategorized. Upon examination of a sample of these 489 unclassified documents, we found no other dominant descriptors relevant for our purposes. Instead, we found descriptors which indicated either educational level or document topic area. For example, 279 were descriptored "Elementary" or "Secondary Education," 185 were descriptored "Vocational." Of the program evaluation methodology collection of 1,734, 1,146 could be subcategorized according to these two types of descriptors.

Our attempt to characterize the 1,734 core collection of program evaluation methodology documents was therefore only partially successful. Only 72 percent of the documents could be placed in subcategories useful in our search, leaving 489 abstracts to be screened by hand. Not all documents in the core collection had been descriptored in sufficient detail for our purposes.

Although we did achieve a partial preliminary screening using ERIC interactive computer searching, it was accomplished only through trial and error. ERIC provides no direct means of characterizing a collection of documents. Even learning the frequency of certain descriptors in a collection has to be done by querying one descriptive at a time. Furthermore, only 98 document manipulation steps are allowed per run, only one document file may be saved at the end of a run, and only 600,000 documents may be inquired about in one run. ERIC presently

contains approximately 316,000 documents, and since learning that 34,949 documents are descriptored Higher Education constitutes an inquiry of 34,949 documents, one cannot even check each document in the system twice. These design constraints greatly hamper attempts to use the ERIC system in an interactive mode on large searches—the searches for which it is most needed.

One additional feature of ERIC should be mentioned. Beginning in September of 1974, all documents entered into ERIC were to be identified by one of 19 "document type" codes such as:

B Books; Monographs, Textbooks; Programmed Texts;

C Curriculum Guides; Teacher-Developed Materials; Laboratory Manuals;

J Journal Articles; Serials; Periodicals; Bulletins; Newsletters; Newspapers;

K Program/Project Descriptions;

R Reports (Research and Technical); (ERIC, 1974)

Out of our 1,734 core documents, 680 documents (EDs) and 232 journal articles (EJs), for a total of 912, were entered after September of 1974. However, document type codes were available for only 676 (74 percent) of these 912 entries. Apparently there is some systematic bias in these omissions, since of the 676 coded, 367 were reports (R) but only 6 were journal articles (J)—a quick check easily verified that far more

than 6 journal articles entered in ERIC since 1974 were included in our core collection. The high percentage of omissions and the probable bias dissuaded us from using the document type codes to characterize our core collection.

In conclusion, our attempts to use ERIC to characterize and initially screen our core literature collection met with limited success. But perhaps we were asking more of the ERIC system than was reasonable. The system was designed primarily for storage and retrieval of fugitive educational documents. We could still hand screen 1,734 abstracts if necessary, assuming the collection really contained the core of the literature we wished to review.

Identifying Literature Using ERIC

How well did our collection of 1,734 abstracts cover the existing literature on program evaluation methodology? We had some doubts about its comprehensiveness, not only because we had already ruled out a lot of apparently related literature, but primarily because one cannot perform hierarchial searches in ERIC. The ERIC Processing Manual warns coders to match the level of the descriptors to the specific level of the document and to avoid descriptors broader or narrower than the scope of the document itself (ERIC, 1974, p. 218-219). Thus a very specific document on the use of case study methods in evaluation will be descriptored with the narrower term "Case Study" rather than the broader term "Evaluation Methods." A search using "Evaluation Methods" will therefore pull only documents descriptored at that broad level and will not pull documents descriptored at the more narrow level of "Case Study," in spite of the fact

that such documents may discuss case studies as an evaluation method. In ERIC, a search employing broad terms will not identify all relevant documents descriptored using more specific terms; i.e., ERIC documents are not descriptored hierarchially.

An external criterion was needed to test the quality of our core literature collection and the recent evaluation bibliography of the Evaluation Center at Western Michigan University was selected (Bunda, 1976). This bibliography contains 603 references on evaluation and is one of the most current and comprehensive bibliographies available. This bibliography is not limited to program evaluation methodology (and we know of no such narrow bibliography currently available) but it is "...unique in its calculated inclusion of fugitive references and a variety of resources" (Bunda, 1976, p. iv). Therefore, more so than some of the previous evaluation bibliographies which focused on the "classics," this bibliography emphasizes fugitive documents just as ERIC does and so should serve as an interesting criterion measure.

Using the Current Index to Journals in Education and the ERIC Ed Accessions File Title Index (through December 1977), each of the 603 references were checked to see if they were contained in ERIC. This information told us the maximum possible overlap between the Western Michigan bibliography of 603 items and our ERIC generated collection of 1,734 items.

The results were startling. Of the 603 references, only 161 (27 percent) were listed in ERIC--less than a third of the Western Michigan bibliography could be reproduced from the ERIC system. Further investigation

indicated why the overlap was so small. Of the 603 references:

73 (12%) were journal articles published prior to 1969--ERIC did not begin including journal articles in its collection until 1969,

83 (14%) were books published prior to 1966--ERIC did not exist until 1966,

12 (2%) were fugitive reports written prior to 1966--ERIC did not exist until 1966.

Therefore, 28 percent of the Western Michigan bibliography was published or written before ERIC was incorporating such documents into its system. But what about the remaining 45 percent of the bibliography which is not in ERIC?

132 (22%) of the references are books or chapters in books published between 1966 and 1976. Because of copyright and reproduction restrictions, commercial publications are not usually made available through ERIC. In fact, ERIC tries to keep the number of documents it cannot reproduce down to 18 percent of its total collection (ERIC, 1974, p. 36). Thus books and other published materials, though current and relevant, will not be found in ERIC. This results in major omissions; for example, neither the AERA nor the CSE evaluation monograph series are referenced in ERIC.

56 (9%) of the omitted Western Michigan references are journal articles published between 1969-1976. Since ERIC includes all major journals, how were these articles missed? The answer is that ERIC in fact does not reference the complete contents of all major journals. The Current Index to Journals in Education July-December 1977 issue lists 739 journals that were

screened during that period. However, only 185 journals (25%) had all of their articles entered into ERIC; for the remaining 554 journals, only selected articles were entered. Most journals are assigned to specific ERIC Clearinghouses which decide how much of a journal's content to enter. Thus if an evaluation article appears in a journal assigned to a Clearinghouse other than ERIC TM, it may not fall within the interests of that Clearinghouse and will either simply not be selected, or at most referred to ERIC TM. An ERIC search, therefore, does not assure one of covering all relevant journal articles.

86 (14%) of the omitted references were fugitive documents and reports written during 1966-1976. These documents should have been entered into ERIC according to the system's guidelines. Either the ERIC TM Clearinghouse never became aware of the existence of these documents or they did not meet ERIC's selection criteria of reproducibility and importance.

The 161 documents of the Western Michigan bibliography which were in ERIC were of the following types:

43 were journal articles published between 1969 and 1976,

6 were books or chapters in books published prior to 1966,

51 were books or chapters in books published between 1966 and 1976,

5 were fugitive documents written prior to 1966, and

56 were fugitive documents written between 1966 and 1976.

To determine the overlap between our 1,734 core documents and the Western Michigan bibliography, the ED or EJ number of each of the 161 references from the bibliography was checked against the 1,734 ED and EJ

numbers of the core collection. The overlap turned out to be an astounding 7 documents of which:

2 were journal articles published 1969-1976,

3 were books or chapters in books published 1966-1976, and

2 were fugitive documents written 1966-1976.

Why were 154 documents of the bibliography not picked up in our search, even though they were in ERIC? An examination of these 154 documents and their descriptors uncovered four general reasons.

First, some items in the bibliography were not directly related to the methodology of program evaluation but focused on such topics as research ethics and standardized testing.

Second, some items related to specific evaluation methodologies such as the Delphi and critical incident techniques. These were not selected in the search because broader descriptors such as "Evaluation Methods" had been used.

Third, some descriptoring, although done in accordance with ERIC guidelines, makes identifying certain documents problematic. For example,

(a) Scriven's chapter "Evaluation Perspectives and Procedures" has "Summative Evaluation" as a major descriptor, while the book it appears in, Popham's Evaluation in Education: Current Applications (1974), does not,

(b) Wright and Worthen's report (1975a) on evaluation contracting has different descriptors than the summary of that report (1975b) which is also in ERIC (the report and summary were descriptored by different coders), and

(c) A design developed by Stufflebeam et al. (1971) for evaluating R&D institutions and programs is not descriptored under "Evaluation Methods", "Methods Research", or "Methods", while a similar design developed by Scriven et al. (1971) is descriptored "Evaluation Methods". These two reports were commissioned by the Office of Education at the same time for the same purpose; they were entered into ERIC by the same coder, but contain different descriptors.

The practice of descriptoring only the document in hand makes it possible that two highly related documents may have different descriptors and not be retrievable under the same search strategy.

Fourth, the assignment of descriptors obviously depends on the background, experience, and point of view of the coder. To see the relevance of certain documents to program evaluation methodology, or any other specialty, requires considerably more familiarity with the field than that possessed by most graduate students who do the majority of Clearinghouse coding (the ERIC Manual does recommend that at least one content specialist participate in the coding of entries).

While the lack of overlap between the Western Michigan bibliography and the 1,734 core program evaluation methodology literature collection can be explained in terms of the deficiency of the bibliography as a proper criterion measure, the design constraints of the ERIC descriptoring system, and human limitations, the trivial overlap is certainly not reassuring.

Conclusions

Large, integrative literature reviews are becoming increasingly needed and increasingly difficult due to the rapidly expanding literature in education. We have attempted to use the ERIC system not only to identify relevant literature but also to screen it. Because of the lack of sufficient depth of descriptoring, it was possible to subcategorize only 72 percent of our initial document collection. ERIC currently provides no means to characterize an existing document collection and even the system's "document type" code facility was available for only three-fourths of the documents it was designed to cover. While initial screening of small document collections is possible, system design constraints prevent complex screening or screening large document files.

ERIC's lack of success in identifying major evaluation documents is even more troublesome, but understandable when one realizes that

(a) no journals before 1968 are in ERIC,

(b) only approximately 25 percent of major education journals are fully covered by ERIC,

(c) few fugitive documents and books prior to 1966 are in ERIC,

(d) few books or other major published materials are in ERIC (maximum of 18% of the system), and

(e) only fugitive materials identified by the specific Clearinghouses of interest are in ERIC.

The added problems of non-hierarchial searching, system descriptoring constraints, and coder limitations forces one to a conservative view of the products of an ERIC search.

The results of the foregoing search can, of course, not be freely generalized to other subsets of literature with ERIC. However, several of the problems we have identified are in fact due to system limitations and will obtain regardless of the descriptors being used.

With the growing literature in education we are forced to use such systems as ERIC, but we must fully understand their limitations. With small, focused searches it may be possible to identify comprehensive sets of documents and review the abstracts by hand. With larger searches, a more sophisticated, automated screening capability is needed than now exists in ERIC. Until such a capability exists, researchers will need to augment ERIC searches with the more traditional literature identification and search procedures. ERIC is currently not an adequate solution to our problems of identifying and characterizing the burgeoning literature in education.

NOTE

[1]I wish to express my sincere thanks to Donna Shaver who collaborated on the computerized ERIC searches and to Denise Smith who assisted with several clerical and editorial tasks.

[2]The searches described here were conducted in mid-May of 1978 and covered all ERIC entries from its inception in 1966 to that time. The searches were conducted using the Lockheed Information System's Dialog System. Although Dialog is the largest information-base access system, an international system accessing approximately 75 data bases, its design is not necessarily representative of other access systems. Thus the computer access design limitations discussed in this paper pertain only to the Dialog System as it is now instituted.

REFERENCES

Bunda, M. A. (Ed.) Evaluation Bibliography, Evaluation Center, Western Michigan University, Kalamazoo, November, 1976.

Educational Resources Information Center, Processing Manual, July 1974. ED 092164.

Glass, G. V , Primary, secondary, and meta-analysis of research. Educational Researcher, 5 (10), 1976, 3-8.

Light, R. J. and Smith, P. V. Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 41(4), 1971, 429-471.

Popham, W. J. (Ed.) Evaluation in education: Current application. Berkeley: McCutchan, 1974. ED 123230.

Scriven, M., et al. An evaluation system for regional labs and R&D centers. Columbus: Ohio State University, 1971. ED 061299.

Stufflebeam, D. L., et al. Design for evaluating R&D Institutions and programs. Columbus: Ohio State University, 1971. ED 061298.

Wright, W. J. and Worthen, B. R. Standards and procedures for development and implementation of an evaluation contract. Portland, Oregon: Northwest Regional Educational Laboratory, 1975(a). ED 127341.

Wright, W. J. and Worthen, B. R. Summary and overview of standards and procedures for evaluation contracting. Portland, Oregon: Northwest Regional Educational Laboratory, 1975(b). ED 127342.